PERSONAL MULTIMEDIA DEVICE
VIDEO FORMAT CONVERSION
ACROSS MULTIPLE VIDEO FORMATS

BACKGROUND OF THE INVENTION

A.   Field of the Invention

[0001]   The present invention relates generally to communications systems, and more particularly, to videoconferencing between entities that use multiple video standards.

B.   Description of Related Art

[0002]   Personal multimedia devices are devices that provide multimedia services to a user. Example of personal multimedia devices include set-top boxes and personal video recording (PVR) devices. A set-top box is a device that enables a television set to receive and decode digital television (DTV) broadcasts. Set-top boxes are frequently used to receive satellite and cable television signals. At a basic level, a set-top box will include an input connection for the television signal broadcast by the satellite or cable company, and an output connection leading to the user's television.

[0003]   Some set-top boxes additionally include more advanced features, such as a network port (e.g., Ethernet) and/or other input output connections, such as a keyboard or a universal serial bus (USB) connection. With these set-top boxes, in addition to simply watching television, users may perform a number of more interactive activities, such as surfing the web and sending email.

[0004]   An additional activity that may be performed in conjunction with the more advanced set-top boxes is videoconferencing. More specifically, a video camera may be connected to the set-top box through a connection such as the USB connection. Audio and video recorded by the video camera may be processed by the

set-top box and then transmitted through the network port to a network, such as the Internet. At the receiving end, another set-top box may receive the video signal from the network, process the video signal into a format compatible with the receiving television, and present the video signal on the receiving television.

[0005] There are situations, however, in which different set-top boxes in the above-described video conferencing scheme are incompatible with one another. Set-top boxes in different regions of the world may use different video formats. Televisions in Europe, for example, typically use the Phase Alternation Line (PAL) analog television display standard while televisions in North America typically use the National Television Systems Committee (NTSC) standard. When attempting to implement a videoconference with televisions using different standards, the received video will not be able to be appropriately displayed on the receiving television.

[0006] Therefore, there is a need in the art to improve video conferencing capabilities of personal multimedia devices such as set-top boxes.

SUMMARY OF THE INVENTION

[0007] A personal multimedia device, as described herein, converts received video signals to a format native to the device. The device may automatically detect a format of the received video signal. The conversion may be performed by adding or removing frames to the received video signal so that the frame rate of the received video signal is close to that of the native format.

[0008] One aspect consistent with the invention is directed to a personal multimedia device that includes a media processing component that increases a frame rate of a received video signal, when the frame rate of the received video signal is less than a frame rate of a native video format of the personal multimedia device by

adding frames to the received video signal where the added frames are based on at least one of the received frames. Further, the media processing component decreases the frame rate of the received video signal when the frame rate of the received video signal is greater than a native frame rate of the personal multimedia device by removing frames from the received video signal.

[0009] Another aspect consistent with the invention is directed to a personal multimedia device that includes a media processing component that detects a frame rate of a received video signal and compares the frame rate to a frame rate native to the personal multimedia device. The media processing component modifies a frame rate of the received video signal when the frame rate of the received video signal is different than the frame rate native to the personal multimedia device by temporally filtering the sequences of frames in the received video signal and also deriving a new set of interpolated frames at the new frame rate.

[0010] Another aspect consistent with the invention is directed to a method for converting a compressed video signal in a first format to a second format. The method includes determining a frame rate of the video signal in the first format based on header information included in the compressed video signal and decoding the compressed video signal to uncompress to video signal. The method further includes increasing the frame rate of the video signal in the first format when the determined frame rate is less than a frame rate of the second format by adding frames to the uncompressed video signal of the first format in which each added frame is based on at least one frame in the uncompressed video signal in the first format. Still further, the method includes decreasing the frame rate of the video signal in the first format when the determined frame rate is greater than the frame rate of the second format by removing frames from the uncompressed video signal in the first format. Finally, the

method includes scaling a frame resolution of each of the frames of the video signal in the first format to adjust a resolution of the video signal in the first format to match the resolution of the video signal in the second format.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011]   The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate the invention and, together with the description, explain the invention. In the drawings,

[0012]   Fig. 1 is a diagram illustrating an exemplary video conferencing system;

[0013]   Fig. 2 is a diagram illustrating functional components of a set-top box;

[0014]   Fig. 3 is a flow chart illustrating operation of set-top box in performing video format conversions consistent with aspects of the invention;

[0015]   Fig. 4 is a diagram illustrating increasing the frames-per-second of a detected video signal;

[0016]   Fig. 5 is a diagram illustrating an alternate implementation for increasing the frames-per-second of a detected video signal;

[0017]   Fig. 6 is a diagram illustrating decreasing the frames-per-second of a detected video signal;

[0018]   Fig. 7 is a diagram illustrating temporal filtering and sharpening; and

[0019]   Fig. 8 is a flow chart illustrating methods performed in a videoconference established consistent with aspects of the invention.

DETAILED DESCRIPTION

[0020]   The following detailed description of the invention refers to the accompanying drawings. The same reference numbers in different drawings may

4

identify the same or similar elements. Also, the following detailed description does not limit the invention. Instead, the scope of the invention is defined by the appended claims and equivalents.

[0021]   A personal multimedia device, as described herein, converts between different video standards to thus facilitate videoconferencing and other streaming video applications across different video transmission standards. The conversion can be performed using standard components of a personal multimedia device such as a set-top box and thus does not add additional cost to the set-top boxes.

VIDEO CONFERENCE OVERVIEW

[0022]   Fig. 1 is a diagram illustrating an exemplary video conferencing system 100. Users 101-1 through 101-N (collectively users 101) may use system 100 to videoconference with one another. Network 110 couples users 101 and 102 with one another. Video streams created by users 101 may be transmitted over network 110. Network 110 may be a private network or a public network such as the Internet.

[0023]   For a particular user 101, system 100 may include a video capture device 120, a personal multimedia device 121, and a television 122. Video capture device 120 may be, for example, a conventional video camera. Television 122 presents the video and audio signals transmitted from user 102 via network 110. Television 122 may alternatively be implemented as other types of multimedia presentation devices, such as a computer monitor and speakers.

[0024]   Personal multimedia device 121 is connected to network 110. Personal multimedia device 121 is also connected to receive the video signals from video capture device 120 and to output television compatible video signals to television 122.

[0025]   Consistent with an aspect of the invention, personal multimedia device 121

may be programmed to convert between different video formats native to different

ones of televisions 122. For example, video capture device 121 of user 101-1 may

output PAL formatted signals and television 122 of user 101-2 may be a NTSC

standard television. In this situation, personal multimedia device 121 of user 101-2

may convert the video signal transmitted over network 110 into an NTSC compatible

television before outputting the signal to television 122 of user 101-2.

[0026]   In one implementation, personal multimedia device 121 is a set-top box.

Accordingly, personal multimedia device 121 may alternatively be referred to herein

as set-top box 121.

SET-TOP BOX OVERVIEW

[0027]   Fig. 2 is a diagram illustrating functional components of a set-top box, such

as set-top box 121. Three functional components are shown in Fig. 2 for simplicity:

channel decoder 210, media processor 220, and input/output interface 230.

[0028]   Video signals may be received by set-top box 121 at channel decoder 210.

Depending on the type of set-top box, channel decoder 210 may include one or more

of satellite transceiver 211, terrestrial decoder 212, and cable transceiver 213.

Satellite transceiver 211 may be designed to receive television signals broadcast from

a satellite. Satellite transceiver 211 may thus include one or more conventional

directional antennas (not shown) capable of transmitting/receiving signals via radio

frequency waves. Satellite transceiver 211 may convert the received signals to digital

signals using conventional techniques and transmit the converted signals to media

processor 220.

[0029]   Terrestrial decoder 212 may include logic for receiving and processing conventional broadcast television signals. Terrestrial decoder 212 may include, for example, an omni-directional antenna (not shown) and circuitry for processing television signals received over the omni-directional antenna. A digital signal resulting from the processing from terrestrial decoder 212 may be output to media processor 220.

[0030]   Cable transceiver 213 receives signals transmitted over a cable system, such as an optical or coaxial cable system, and converts the signals into a digital signal. Media processor 220 receives the signals from channel decoder 210 and processes the signals into a form suitable for viewing on television 122. Media processor 220 may include a number of functional components, such as MPEG decoder/encoder 221, graphics processor(s) 222, and video encoders/decoders 223.

[0031]   MPEG decoder/encoder 221 may perform encoding and decoding functions relating to the Moving Picture Experts Group (MPEG) family of standards. The MPEG standards are used for coding audio-visual information in a digital compressed form. Because MPEG video streams are compressed, they can be transmitted with significantly less bandwidth than uncompressed streams. Multiple versions of the MPEG standard exist, such as the MPEG-2 and MPEG-4 standards. MPEG decoders/encoders are known in the art and will not be described further herein.

[0032]   Graphics processor(s) 222 perform image-processing functions on an uncompressed image. Graphics processor(s) 222 may, for example, scale an image, manipulate the color in the image, or superimpose text or graphics on an image. Graphics processors are generally known in the art and are frequently implemented using custom semiconductor chips designed to efficiently perform graphic functions.

[0033] Video decoder/encoder 223 may be configured to convert uncompressed video streams into analog television signals. Video decoder/encoder 223 can be, for example, a PAL or NTSC decoder/encoder that contains logic to convert video signals to one or both of these standards. Video decoders/encoders are generally well known in the art.

[0034] Input/output interface 230 may include logic for connecting additional devices to set-top box 121. Input/output interface 230 may include network ports, such as an Ethernet port, keyboard and mouse ports, USB ports, etc. Through input/output interface 230, set-top box 121 can provide personal computer-like functions. In particular, through a network port such as an Ethernet port, set-top box 121 may connect and communicate with other devices, including other set-top boxes coupled to network 110. Two set-top boxes may, for example, exchange MPEG video streams over network 110.

[0035] The components illustrated in Fig. 2 are functional components useful in explaining the present invention. Actual set-top boxes may contain many additional components. Also, in actual set-top boxes, multiple ones of the components shown in Fig. 2 may be implemented as a single physical device or conversely, multiple physical devices may be used to implement one of the functional components shown in Fig. 2. It should also be understood that the components illustrated in Fig. 2 may be implemented in hardware, software, or any combination of hardware and software.

[0036] In alternate implementations, instead of using the MPEG family of compression standards, other compression standards, such as H.261, H.263, or H.264 may be used. In general, the H.261, H.263, and H.264 compression standards use best effort to achieve a frame rate of 15 or 30 frames-per-second.

SET-TOP BOX OPERATION

[0037]　As discussed above, video or television signals may be implemented in one of a number of different formats. The PAL standard, for example, has a line resolution of 625 lines per frame and is displayed at 25 frames-per-second (fps). The NTSC standard has a line resolution of 525 lines per frame and is displayed at 29.97 fps.

[0038]　Set-top boxes 121 will generally be programmed to output a "native" format that is common to the region in which the set-top box is used. Consistent with an aspect of the invention, media processor 220 converts between incoming video formats and the native format of set-top box 121. The conversion is performed using hardware that is conventionally available in many set-top boxes, and can thus be implemented in set-top box 121 with relatively little incremental cost. Fig. 3 is a flow chart illustrating operation of set-top box 121 in performing such conversions.

[0039]　Incoming video signals may be processed by channel decoder 210 (Act 301). The output of channel decoder 210 may be a digital version of the video signal. The video signal may be compressed using a video compression standard such as MPEG-2 or MPEG-4. The MPEG bit stream may be decoded by MPEG decoder/encoder 221 to produce a non-compressed video signal in a format such as PAL or NTSC. Alternatively, the incoming video signals may be received as a digital video bit stream (e.g., as an MPEG bit stream) over network 110 by input/output interface 230. In this situation, the digital video stream may be transmitted directly to media processor 220.

[0040]　Information relating to the format of a video signal, such as the frame rate, may be embedded in fields associated with the MPEG bit stream. Media processor 220 may examine fields in the MPEG video stream to determine the format, such as the frame rate and resolution of the video signal. In one implementation, the video

signal is compressed using either the MPEG-2 or MPEG-4 standard. Media processor 220 may examine the MPEG bit stream to determine the appropriate MPEG version (Act 302). This information is contained in header fields present in the MPEG bit stream.

[0041] Based on the MPEG version determined in Act 302, media processor 220 determines the format of the video signal (Acts 303 and 304). For MPEG-2 video streams, the frame rate is described in a four-bit "frame rate" field. Also, under MPEG-2, a three-bit "video format" field describes the video format encoded in the MPEG bit stream. Table I and II, below, describe the meaning for various values in the "frame rate" and "video format" fields under MPEG-2.

TABLE I—MPEG-2 Frame Rate Field	
Frame Rate Code	Frame Rate Value
0000	Forbidden
0001	24.000 ÷ 1.001 (23.976)
0010	24
0011	25
0100	30.000 ÷ 1.001 (29.97)
0101	30
0110	50
0111	60.000 ÷ 1.001 (59.94)
1000	60
1001	Reserved
...	...
1111	Reserved

TABLE II—MPEG-2 Video Format Field	
Video Format Code	Meaning
000	Component
001	PAL
010	NTSC
011	SECAM (Systeme Electronique Couleur Avec Memoire)
100	MAC
101	Unspecified video format
110	Reserved
111	Reserved

[0042]   Thus, for MPEG-2 bit streams, media processor 220 may use either the frame rate field to infer the format of the video stream or the video format field to directly determine the format of the video stream (Act 303). As an example of inferring the format of a video stream, consider the situation in which the frame rate is 29.97 fps. This frame rate corresponds to an NTSC video signal.

[0043]   MPEG-4 video streams include a video format field similar to the video format field contained in MPEG-2 video streams. Accordingly, for MPEG-4 video streams, the video format field can be used to determine the format, and hence the frame rate, of the video stream (Act 304).

[0044]   If the frame rate of the format detected in Acts 303 and 304 is the same as the native format of set-top box 121, the received video stream can be played back by set-top box 121 without conversion (Acts 305 and 306). In this situation, MPEG decoder/encoder 221 may decode the MPEG video stream to an uncompressed

version, which may then be converted directly to the appropriate analog output format by video decoder/encoder 223.

[0045] When, however, the frame rate of the format detected in Acts 303 and 304 is different than the native format of set-top box 121, media processor 220 determines whether the native format has more frames-per-second than the native format (Act 307). If the native format has more frames-per-second than the detected format, media processing component 220 increases the frames-per-second of the detected video signal (Act 308).

[0046] Fig. 4 is a diagram illustrating increasing the frames-per-second of the detected video signal as performed in Act 308. The video signal may first be decoded by MPEG decoder/encoder 221 to obtain an uncompressed version of the video signal. Assume that for this example, the detected video format is PAL (25 fps) and the native video format is NTSC (29.97 fps).

[0047] In Fig. 4, a PAL video signal 410 is represented as a number of frames 411-1 through 411-5. Although only five frames are shown in Fig. 4, in practice, the process shown for the five frames in Fig. 4 may be repeated for each set of five frames in the video signal. PAL video signals are in an interlaced format. In an interlaced video signal, each frame is rendered in two successive sweeps. Thus, for example, with a PAL frame that includes 625 lines, the odd lines of the frame may be rendered first followed by the even lines. In this manner, when rendering the frame, the television interlaces the odd and even portions of the frame. The odd and even interlaced portions of frames 411 are indicated in Fig. 4 with "A" for the odd fields and "B" for the even fields.

[0048] The target video format, an NTSC signal, is illustrated in Fig. 4 as video signal 420. Video signal 420 includes six frames 421-1 through 421-6. Thus, as

shown in this figure, the five frames in PAL video signal 410 are converted into

NTSC video signal 420 containing six NTSC frames 421.

[0049]   The first frame of NTSC video signal 420, frame 421-1, is formed from

the odd and even portions of the first frame of PAL video signal 410, frame 411-1.

That is, frame 421-1 is essentially the same as frame 411-1, although frame 421-1 is

associated with a shorter time duration. Similarly, frame 421-2 is formed from the

odd and even portions of frame 411-2. Frame 421-3, however, is formed from the

odd portion of frame 411-2 and the even portion of 411-3. Frame 421-4 is formed

from the odd portion of frame 411-3 and the even portion of 411-4. Frame 421-5 is

formed from the odd and even portions of frame 411-4. Frame 421-6 is formed from

the odd and even portions of frame 411-5.

[0050]   In the operations shown in Fig. 4, five PAL formatted frames are

converted to six NTSC formatted frames. A pure five-to-six conversion would

actually produce 30.00 NTSC fps instead of the required 29.97 fps. To correct this,

MPEG decoder/encoder 221 may additionally multiply the converted 30.00 fps signal

by 0.999 to obtain a 29.97 fps video signal. One way of physically implementing the

multiply operation is by dropping a frame every 1001 frames.

[0051]   Fig. 5 is a diagram illustrating another implementation of Act 308. In this

example, a FILM formatted video signal is converted to an NTSC formatted video

signal. The FILM format is a 24 fps format compared to the 29.97 fps format of

NTSC.

[0052]   FILM video signal 510 is represented as a number of frames 511-1 through

511-4. Although only four frames are shown in Fig. 5, in practice, the process shown

for the four frames in Fig. 5 may be repeated for each set of four frames in the video

signal. The target video format, an NTSC signal, is illustrated as video signal 520. Video signal 520 includes frames 521-1 through 521-5.

[0053]   As shown in Fig. 5, the first frame of NTSC video signal 520, frame 521-1, is formed from the odd and even portions of the first frame of FILM video signal 510, frame 511-1. Similarly, frame 521-2 is formed from the odd and even portions of frame 511-2. Frame 521-3, however, is formed from the odd portion of frame 511-2 and the even portion of 511-3. Frame 521-4 is formed from the odd and even portions of frame 511-3. Frame 521-5 is formed from the odd and even portions of frame 511-4.

[0054]   Referring back to Fig. 3, when the frame rate of the format detected in Act 307 is different than the native format of set-top box 121, and the native format has less frames-per-second than the detected format, media processing component 220 decreases the frames-per-second of the detected video signal (Act 309).

[0055]   Fig. 6 is a diagram illustrating decreasing the frames-per-second of the detected video signal as performed in Act 309. Assume that for this example, the detected video format is NTSC (29.97 fps) and the native video format is PAL (25 fps). In this situation, media processor 220 occasionally drops frames to reduce the frames-per-second. For most applications, an occasional dropped frame will not be noticed when converting from NTSC to PAL video formats.

[0056]   In Fig. 6, a NTSC video signal 610 is represented as a number of frames 611-1 through 611-6. The target video format, a PAL signal, is illustrated in Fig. 6 as video signal 620. Video signal 620 includes five frames 621-1 through 621-5. Thus, as shown in this figure, the six frames in NTSC video signal 610 are converted into PAL video signal 620 containing five PAL frames 620. The conversion is formed by

copying, on a one-to-one basis, frames 611-1 to 611-5 to frames 621-1 to 621-5, respectively. Frame 611-6 is dropped.

[0057]   A straight six-to-five frame conversion from NTSC would actually yield a video signal at 24.975 fps. Accordingly, MPEG decoder/encoder 221 may additionally divide the PAL video signal 620 by 0.999 to achieve 25 fps. In practice, the divide operation may be performed by occasionally adding a duplicate frame to the signal.

[0058]   Figs. 4-6 illustrate frame rate conversion for video signals having interlaced formats. For video formats having progressive (non-interlaced) frames, increasing or decreasing the fps of the video signal reduces to a simple mathematical relation of dropping/repeating X frame(s) every Y frames(s). Thus, for example, if a first progressive video format having 10 fps is to be converted to a second video format having 15 fps, every other frame in the first video format could be repeated to form the second video format.

[0059]   The techniques shown in Figs. 4-6 do not impact lip-sync because the Audio Presentation Time Stamp (PTS) does not change. Accordingly, these techniques do not require modification of the audio stream corresponding to the video signal.

[0060]   Referring back to Fig. 3, graphics processor(s) 222 may rescale the picture size by scaling the resolution of each of the frames. In particular, if the detected frame size and the native frame size are not the same, (Act 310), graphic(s) processor(s) 222 may rescale the frame size so that the converted video signal has the correct frame size (Act 311). In the case of PAL to NTSC conversion, graphics processor(s) 222 scales the vertical scale by a factor of 0.84. Conversely, for NTSC to PAL conversion, graphics processor(s) 222 stretches the vertical scale by a factor

of 1.19 (1/0.84). Alternatively, black bars can be inserted if it is desirable to maintain the aspect ratio.

[0061]   Video decoder/encoder 223 may next process the converted video signal into an analog format appropriate for television 122 (Act 312). Video decoder/encoder 223 may, for example, add the frame timing information required by the particular video format.

[0062]   The above-discussed implementations perform frame rate conversion using relatively simple frame add/drop techniques for increasing/decreasing frame rate. In other implementations, set-top box 121 may implement more complicated image conversion techniques. For example, set-top box may perform temporal filtering and image sharpening to more optimally convert video signals.

[0063]   Fig. 7 is a diagram illustrating interpolation through temporal filtering in additional detail for an exemplary NTSC to PAL conversion. NTSC video signal 710 include six frames 711-1 through 711-6. Instead of simply dropping every sixth frame as described above with reference to Fig. 6, frames may be temporarily filtered and sharpened in a more optimal (although potentially more processor/memory resource intensive) manner. Thus, target frame 721-2, for example, may be formed as a combination of the information in frames 711-3 and 711-4, or even as a combination of the information in frames 711-2 through 711-5.

VIDEO CONFERENCING

[0064]   As previously mentioned, one application of the above-described video format conversion performed by set-top box 121 is videoconferencing. Fig. 8 is a flow chart illustrating methods performed in a videoconference established in a manner consistent with aspects of the invention.

[0065]   To begin, two or more set-top boxes may communicate with one another over network 110 to form a network connection (Act 801). For example, user 101-1 may form a video conference connection with users 101-2 and 101-3. In some implementations, devices other than set-top boxes 121 may also participate in the videoconference. For example, a set-top box may communicate with a personal computer to establish a videoconference.

[0066]   During the videoconference, data received the video capture device 120 associated with set-top boxes 121 is transmitted over network 110 (Act 802). The video will generally be transmitted at the frame rate and resolution native to the transmitting set-top box or video capture device. Set-top boxes 121 also receive and display video signals over network 110 (Act 803). Video signals that are not in the native format of the set-top box are converted as described above. As described above, a video signal can be converted to the appropriate video signal format on-the-fly, without having to pre-negotiate a particular video standard to use in the video conference. Thus, each of the participants of the video conference could potentially be using a different native video standard, yet the video conference could still proceed without having to first negotiate a common standard. Relative to the centralized format decisions common in broadcast television, the format decision is made in a distributed manner.

[0067]   In the videoconferencing application described above, a standard set-top box may be used to implement videoconferencing with videoconference partners that transmit video in a format different than the native format of the set-top box.

CONCLUSION

[0068]   Techniques described herein provide for conversion between video

formats at a personal multimedia device using existing hardware in the personal

multimedia device.

[0069]   It will be apparent to one of ordinary skill in the art that aspects of the

invention, as described above, may be implemented in many different forms of

software, firmware, and hardware in the implementations illustrated in the figures.

The actual software code or specialized control hardware used to implement aspects

consistent with the present invention is not limiting of the present invention. Thus,

the operation and behavior of the aspects were described without reference to the

specific software code or hardware logic. It should be understood that a person of

ordinary skill in the art would be able to design software and control hardware to

implement the aspects of the present invention based on the description herein.

[0070]   The foregoing description of preferred embodiments of the present

invention provides illustration and description, but is not intended to be exhaustive or

to limit the invention to the precise form disclosed. Modifications and variations are

possible in light of the above teachings or may be acquired from practice of the

invention. For example, devices other than traditional set-top boxes may be used to

convert between video formats. Video game consoles, for example, may be used to

implement video conferences consistent with aspects of the invention.

[0071]   No element, act, or instruction used in the description of the present

application should be construed as critical or essential to the invention unless

explicitly described as such. Also, as used herein, the article "a" is intended to

include one or more items. Where only one item is intended, the term "one" or

similar language is used.

[0072]   The scope of the invention is defined by the claims and their equivalents.